(19)日本国特許庁（ＪＰ）　(12) 公開特許公報（A）　(11)特許出願公開番号

特開平7－219929

(43)公開日　平成７年(1995)８月18日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　17/18					
Ｇ０５Ｂ　19/418					
			Ｇ０６Ｆ　15/ 36	Ｚ	
		7531－3Ｈ	Ｇ０５Ｂ　15/ 02	Ｓ	

審査請求　未請求　請求項の数18　ＯＬ　（全 32 頁）

(21)出願番号　特願平6－8871

(22)出願日　平成６年(1994)１月28日

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号

(72)発明者　上田　太一郎
横浜市戸塚区川上町87番地１　三菱電機東部コンピュータシステム株式会社横浜システムセンター内

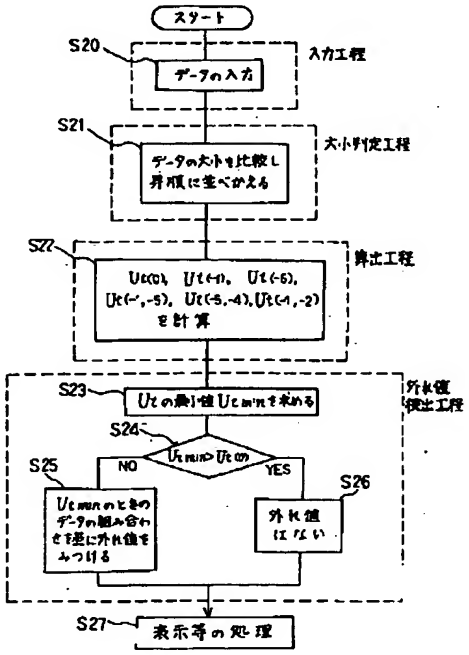
(74)代理人　弁理士　高田　守

(54)【発明の名称】　外れ値検出方法及びデータ処理装置

(57)【要約】

【目的】　得られたデータ（収量、反応量）の外れ値を求める。

【構成】　入力工程でデータを入力し、大小判定工程で昇順に並べる。次に、算出工程で一番小さい値を除いた検出統計量を計算する。一番大きい値を除いた検出統計量を計算する。一番小さい値と一番大きい値を除いた検出統計量を計算する。以下同様にして考えられる組み合せのデータから検出統計量を計算する。またデータを除かないときの検出統計量も計算する。外れ値検出工程で、以上の検出統計量が最小となるデータの組み合せを見つけ、その除いた値を外れ値とする。データを除かないときの検出統計量が最小となったとすれば外れ値は存在しない。



1

【特許請求の範囲】

【請求項１】　以下の工程を有する外れ値検出方法（ａ）Ｎ個（Ｎ≧３）の値を入力する入力工程、（ｂ）上記入力工程により入力したＮ個の値の大小関係を判定する大小判定工程、（ｃ）上記大小判定工程により判定された大小関係に基づき、Ｎ個の値の組み合せ及び外れ値の候補を除いたＮ個未満の値の組み合せを求め、求めた組み合せに対して所定の計算式を用いて検出統計量を算出する算出工程、（ｄ）上記算出工程により算出された検出統計量に基づいて、外れ値を検出する外れ検出工程。

【請求項２】　上記算出工程は、ｓ個以内の外れ値を検出する場合、大小判定工程により判定された大小関係上連続するｎ個（ｎ＝Ｎ－ｓ）以上の値の複数の組み合せを用いて検出統計量を算出することを特徴とする請求項１記載の外れ値検出方法。

【請求項３】　上記外れ値検出工程は、Ｎ個未満の値の組み合せから求めた検出統計量の中で最小のものを選択する最小値選択工程と、選択された最小値がＮ個の値の組み合せから求めた検出統計量よりも小さい場合に、その選択された最小値を算出した組み合せに含まれていなかった値を外れ値とする外れ値判定工程を備えたことを特徴とする請求項１又は２記載の外れ値検出方法。

【請求項４】　上記計算式は、外れ値の候補が除かれると小さくなる傾向にある第１の項目と、外れ値の候補が除かれると大きくなる第２の項目とを有し、上記算出工程は、第１と第２の項目の値を算出し両者の和により検出統計量を求めることを特徴とする請求項１、２又は３記載の外れ値検出方法。

【請求項５】　上記計算式は、更に、第１と第２の項目以外に、第１と第２の項目を補正する補正項を有し、上記算出工程は、第１と第２と第３の項目の値を算出し、３者の和により検出統計量を求めることを特徴とする請求項４記載の外れ値検出方法。

【請求項６】　上記第１の項目は、検出統計量を求めるＮ個未満の値の分散を用いていることを特徴とする請求項４又は５記載の外れ値検出方法。

【請求項７】　上記第２の項目は、検出統計量を求める場合の外れ値の候補の個数を用いていることを特徴とする請求項４又は５記載の外れ値検出方法。

【請求項８】　上記第２の項目は、外れ値の候補の個数に対して所定の係数を乗算したものを用いることを特徴とする請求項７記載の外れ値検出方法。

【請求項９】　上記計算式は、検出統計量を求めるＮ個未満の値の分散と分散に対する係数を有しており、上記算出工程は、分散と係数の乗算により検出統計量を求めることを特徴とする１、２又は３記載の外れ値検出方法。

【請求項１０】　上記計算式は、回帰分析の変数選択基準を基礎にして作成されることを特徴とする請求項１～８又は９記載の外れ値検出方法。

2

【請求項１１】　上記外れ値検出方法は、更に、入力工程と大小判定工程の間に、入力した値を加工する加工工程を備えたことを特徴とする請求項１記載の外れ値検出方法。

【請求項１２】　上記加工工程は、入力工程により入力された時間に依存する値を時間に依存しない値に加工することを特徴とする請求項１１記載の外れ値検出方法。

【請求項１３】　上記加工工程は、入力工程により入力された値からテコ比を計算することを特徴とする請求項１１記載の外れ値検出方法。

【請求項１４】　上記加工工程は、入力工程により入力された値から回帰分析モデルのデータを計算することを特徴とする請求項１１記載の外れ値検出方法。

【請求項１５】　上記加工工程は、入力工程により入力された値から正準相関分析モデルのデータを計算することを特徴とする請求項１１記載の外れ値検出方法。

【請求項１６】　上記加工工程は、入力工程により入力された値が複数のグループに分類されていて複数の要因により判別分析を行う場合に、各グループの判別関数値を計算することを特徴とする請求項１１記載の外れ値検出方法。

【請求項１７】　上記請求項１～１５又は１６記載の外れ値検出方法を実行して外れ値を検出する外れ値検出手段と、Ｎ個の値を計測して外れ地検出手段に入力する計測手段と、外れ値検出手段により検出された外れ値を知らせる出力手段を備えたデータ処理装置。

【請求項１８】　上記データ処理装置は、更に、外れ値検出手段により検出された外れ値を除いた残りの値を用いて所定の処理を実行するデータ処理手段を備えたことを特徴とする請求項１７記載のデータ処理装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は生産工程、品質管理、研究開発、品質改良などにおけるデータの外れ値を検出する方法及びその方法を利用した装置に関するものである。

【０００２】

【従来の技術】例えば、製品の性能バラ付きを測定する場合、あるいは、電力メータや水道メータ等の検針を行う場合、更には実験データを測定する場合に、得られたデータの中に規格外れの性能を示すデータや、異常な測定値を示すデータが存在する。このように、規格外れのデータや、異常値は測定環境や測定装置自身から生ずる不適切なデータであることが多い。このような不適切なデータを、ここでは以下外れ値と呼ぶことにする。外れ値は、本来測定されるべき値ではないため、前述したような各種データから外れ値を検出し、取り除く手法が従来から考えられてきている。データから外れ値を検出する方法は、従来から統計手法に基ずくものがある。外れ

3

値とは極端に大きなあるいは小さい値をとるデータのことである。例えば、５．７１、６．５７、７．２９、８．０６、１０．００、１５．００を考える。プロットすると図１２のようになる。図１２を見ると１５．００は外れ値のようである。統計手法では外れ値が１個として、１個の時の外れ値を検定する計算式を用いる。２個の時は２個用の計算式を用いる。統計的検定であるから予め危険率（有意水準）を決めておく必要がある。危険率としては伝統的に５％あるいは１％を用いている。危険率５％とは、統計的検定により外れ値と判断を下す時誤る確率が５％であることを示す。計算式に対応した５％あるいは１％の数表があり、実データで計算した値と数表とを比較して大ならば外れ値とする。ただし危険率としては、５％ある。または１％あるということになる。

【０００３】このように従来の統計手法では外れ値の個数が１個の時、２個の時、３個の時によって計算法が異なったり、危険率（有意水準とも呼ばれる）の違い（５％、１％等）により結論が異なる（５％の時外れ値と結論しても１％の時は外れ値とはいえない等）問題点がある。また大きな値の外れ値、小さな値の外れ値により計算法が異なる。統計手法であるから５％、１％の数表も必要である。

【０００４】具体的には、図１３及び図１４を用いて説明する。図１３で１は情報処理装置、２はコンピュータ（ＦＤＤ付）、３はディスプレイ・ユニット、４はプリンタ、５はキーボード、６はフロッピーディスクである。プログラム・ルーチンが記憶されたフロッピーディスク６をコンピュータ（ＦＤＤ付）２に挿入し、オペレ*

4

*ーション・ソフトを駆動して、情報処理装置１をスタートさせる。フロッピーディスク６からプログラム・ルーチンがロードされ入力待状態となる。

【０００５】図１４は従来例の説明のためのフローチャートである。ステップ１は、キーボード５からデータを連続的に入力する段階である。ステップ２では、外れ値の個数を入力し、１ならば、ステップ３、ステップ４で、小さい値又は大きい値を外れ値とした統計量をそれぞれ求める。なぜ別々に求めるかと言えば、小さい値と大きい値では計算方式が異なるからである。また、外れ値の個数が２個の場合は、ステップ５からステップ７で、小さい値を２個外れ値とした場合、小さい値と大きい値を１個ずつ外れ値とした場合、大きい値を２個外れ値とした場合にそれぞれ別の計算方式で統計量を求める。

【０００６】ステップ８では、数表を見て上記ステップで求めた計算値と数表にある有意点の大小比較をする。ステップ９では、有意点より計算値の方が大きい場合外れ値と認識する。ステップ１０は、計算値の方が小さい場合外れ値としない。ステップ１１では、結果の表示等をする。図１４は危険率（有意水準）が５％の場合であるが、１％の場合なら１％の数表が必要となる。

【０００７】従来の統計手法で外れ値を検出する例を示す。データとして、

５．７１、６．５７、７．２９、８．０６、１０．００、１５．００

とする。Ｇｒｕｂｂｓ検定量の式、数１を用いる。

【０００８】

【数１】

$$T_i = (x_i - \bar{x}) / S$$

$$\text{ただし } S = \sum (x_i - \bar{x}) / (n-1), \qquad n：\text{サンプル数}$$

【０００９】１５．００が外れ値と考えられるので、Ｔｉの最大値ｍａｘＴｉ（ｉ＝１、２、・・・、ｎ）を求めて数表に載っている値と比較する。

ｍａｘＴｉ＝１．８４１

となった。Ｇｒｕｂｂｓの数表、表１を見るとサンプル数ｎ＝６の時、かつ、危険率５％の時１．８２、サンプル数ｎ＝６、かつ、危険率１％の時１．９４である。よって、

ｍａｘＴｉ＞１．８２、ｍａｘＴｉ＜１．９４

である。従って、１５．００は危険率５％で外れ値といえる。危険率１％では外れ値といえない。このように危険率の違いにより結論が異なってくる。

【００１０】

【表１】

Ｇｒｕｂｂｓの数表の一部

ｎ	５％	１％
４	１．４６	１．４９
５	１．６７	１．７５
６	１．８２	１．９４
７	１．９４	２．１０

【００１１】（Ｖｉｃ Ｂａｒｎｅｔｔ，Ｔｏｂｙ Ｌｅｗｉｓ（１９７８）：「Ｏｕｔｌｉｅｒｓ ｉｎ Ｓｔａｔｉｓｔｉｃａｌ Ｄａｔａ」，Ｊｏｈｎ Ｗｉｌｅｙ＆ Ｓｏｎｓ．ｐ．２９８から一部引用）

【００１２】次に、マスク効果の例をあげる。データとして５．７１、６．５７、７．２９、８．０６、１４．８０、１５．００とする。このデータではｍａｘＴｉ＝１．２９となる。ｍａｘＴｉ＜１．８２である。従って外れ値はないことになる。これはマスク効果といって、

5

上のように外れ値の候補が１４．８０と１５．００の２つある場合、従来方式では外れ値を１つとして検定すると必ずしも外れ値を検出しない例である。

【００１３】

【発明が解決しようとする課題】以上説明したように、従来のものでは外れ値の個数により計算方式が異なる。また、外れ値の性格（大きい方の外れ値か小さい方の外れ値か）により計算方式が異なる（Ｖｉｃ Ｂａｒｎｅｔｔ，Ｔｏｂｙ Ｌｅｗｉｓ（１９７８）：「Ｏｕｔｌｉｅｒｓ ｉｎ Ｓｔａｔｉｓｔｉｃａｌ Ｄａｔａ」，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓには４０種以上の計算式が成っている）という問題点があった。また、計算値と数表の大小比較が必要である。また、危険率の違い（５％、１％等）により結論が異なるという問題点があった。また、マスク効果といって例えば外れ値の候補が２つある場合、従来方式では外れ値を１つとして検定すると必ずしも外れ値を検出しないという問題点があった。

【００１４】この発明は、以上のような問題点を解決するためになされたものであり、従来のような数表を用いることなく、また、外れ値の個数や外れ値の性格により計算方式を変える必要がない外れ値検出方法を得ることを目的とする。また、マスク効果を回避することができる外れ値検出方法を得ることを目的とする。また、外れ値を検出する場合にできるだけ計算過程が簡単で、且つ、計算量も少なくて済む外れ値検出方法を得ることを目的とする。更には、これらの外れ値検出方法を利用したデータ処理装置を提供することを目的とする。

【００１５】

【課題を解決するための手段】この発明に係る外れ値検出方法は、以下の工程を有する。

（ａ）Ｎ個（Ｎ≧３）の値を入力する入力工程、（ｂ）上記入力工程により入力したＮ個の値の大小関係を判定する大小判定工程、（ｃ）上記大小判定工程により判定された大小関係に基づき、Ｎ個の値の組み合せ及び外れ値の候補を除いたＮ個未満の値の組み合せを求め、求めた組み合せに対して所定の計算式を用いて検出統計量を算出する算出工程、（ｄ）上記算出工程により算出された検出等計量に基づいて、外れ値を検出する外れ検出工程。

【００１６】上記算出工程は、ｓ個以内の外れ値を検出する場合、大小判定工程により判定された大小関係上連続するｎ個（ｎ＝Ｎ－ｓ）以上の値の組み合せを複数作成し、これらの組み合せを用いて検出統計量を算出することを特徴とする。

【００１７】上記外れ値検出工程は、Ｎ個未満の値の組み合せから求めた検出統計量の中で最小のものを選択する最小値選択工程と、選択された最小値がＮ個の値の組み合せから求めた検出統計量よりも小さい場合に、その選択された最小値を算出した組み合せに含まれていなかった値を外れ値とする外れ値判定工程を備えたことを特徴とする。

6

【００１８】上記計算式は、外れ値の候補が除かれると小さくなる傾向にある第１の項目と、外れ値の候補が除かれると大きくなる第２の項目とを有し、上記算出工程は、第１と第２の項目の値を算出し両者の和により検出統計量を求めることを特徴とする。

【００１９】上記計算式は、更に、第１と第２の項目以外に、第１と第２の項目を補正する補正項を有し、上記算出工程は、第１と第２と第３の項目の値を算出し、３者の和により検出統計量を求めることを特徴とする。

【００２０】上記第１の項目は、検出統計量を求めるＮ個未満の値の分散を用いていることを特徴とする。

【００２１】上記第２の項目は、検出統計量を求める場合の外れ値の候補の個数を用いていることを特徴とする。

【００２２】上記第２の項目は、外れ値の候補の個数に対して所定の係数を乗算したものを用いることを特徴とする。

【００２３】上記計算式は、検出統計量を求めるＮ個未満の値の分散と分散に対する係数を有しており、上記算出工程は、分散と係数の乗算により検出統計量を求めることを特徴とする。

【００２４】上記計算式は、回帰分析の変数選択基準を基礎にして作成されることを特徴とする。

【００２５】上記外れ値検出方法は、更に、入力工程と大小判定工程の間に、入力した値を加工する加工工程を備えたことを特徴とする。

【００２６】上記加工工程は、入力工程により入力された時間に依存する値を時間に依存しない値に加工することを特徴とする。

【００２７】上記加工工程は、入力工程により入力された値からテコ比を計算することを特徴とする。

【００２８】上記加工工程は、入力工程により入力された値から回帰分析モデルのデータを計算することを特徴とする。

【００２９】上記加工工程は、入力工程により入力された値から正準相関分析モデルのデータを計算することを特徴とする。

【００３０】上記加工工程は、入力工程により入力された値が複数のグループに分類されていて複数の要因により判別分析を行う場合に、各グループの判別関数値を計算することを特徴とする。

【００３１】また、この発明に係るデータ処理装置は、外れ値検出方法を実行して外れ値を検出する外れ値検出手段と、Ｎ個の値を計測して外れ地検出手段に入力する計測手段と、外れ値検出手段により検出された外れ値を知らせる出力手段を備える。

【００３２】上記データ処理装置は、更に、外れ値検出手段により検出された外れ値を除いた残りの値を用いて

所定の処理を実行するデータ処理手段を備えたことを特徴とする。

【００３３】

【作用】第１の発明においては、入力工程により、Ｎ個の値が入力されると、大小判定工程により値の大小関係を判定し、大きい方の値又は小さい方の値のいくつかを外れ値の候補とする。算出工程は、まずＮ個の値の組み合せ及び外れ値の候補を除いたＮ個未満の値の組み合せを求め、次に求めた組み合せそれぞれに対して所定の計算式を用いて検出統計量を算出する。外れ値検出工程は、算出された検出統計量に基づいて外れ値を検出する。

【００３４】第２の発明における算出工程は、大小判定工程により判定された値の大小に基づき、ｎ個（ｎ＝Ｎ－ｓ）以上の連続する値の組み合せを用いて、所定の計算式により検出統計量を計算する。例えば、入力工程により５個（Ｎ＝５）が入力され、最大２個（ｓ＝２）の外れ値を検出しようとする場合、大きい方から３個の入力値を用いて１つの組み合せを作成する。また、大きい方から４個の入力値を用いて別な組み合せを作成する。また、最大値と最小値を除いた中間の値３個を用いて１つの組み合せを作成する。また、小さい方の入力値３個及び小さい方の入力値４個を用いてそれぞれ組み合せを作成する。

【００３５】第３の発明における外れ値検出工程は、まず、算出工程により算出されたＮ個未満の値の組み合せの検出統計量の中で最小のものを選択する。次に、Ｎ個の値の組み合せから求めた検出統計量と、選択された最小値を比較し最小値の方が小さい場合、その選択された最小値を算出した組み合せに含まれていなかった値を外れ値とする。また、最小値の方が大きい場合、外れ値は無しと判定する。

【００３６】第４の発明における計算式は、外れ値の候補が除かれると小さくなる傾向にある第１の項目と、外れ値の候補が除かれる時に大きくなる第２の項目を有し、両者の和により検出統計量を求める。この計算式により、外れ値がある場合最も外れた値が除かれると検出統計量が最小となる。

【００３７】第５の発明における計算式は、上記第１と第２の項目に加えて、第３の項目を持つ。この第３の項目は、上記第１と第２の項目を補正する補正項目である。第１項目、第２項目、第３項目を加算して検出統計量を求める。

【００３８】第６の発明における計算式は、第１項目に検出統計量を求めるＮ個未満の値の分散を含んでいる。従って、最も外れた値が除かれると分散の値が小さくなり、第１の項目の値が小さくなる。

【００３９】第７の発明における計算式は、第２項目に検出統計量を求める場合の外れ値の候補の個数を含んでいる。従って、外れ値の数を多く検出しようとすると、

第２の項目の値が大きくなる。

【００４０】第８の発明における計算式は、第２の項目に外れ値の候補の個数に対して所定の係数を乗算したものを用いる。

【００４１】第９の発明における計算式は、検出統計量を求めるＮ個未満の値の分散に係数を乗算して検出統計量を求める。

【００４２】第１０の発明における計算式は、回帰分析の変数選択基準を基礎にして検出統計量を求める計算式を作成する。

【００４３】第１１の発明においては、加工工程により、入力工程により入力された値を、検出統計量を求めることができるデータに変換することができるため、様々な種類のデータを入力することができる。

【００４４】第１２の発明においては、入力工程により入力された値から例えば時間に比例して増加、あるいは、減少する傾向を補正して時間に依存しない値に加工する。そして、補正された値から検出統計量を算出し、外れ値を求めることができる。

【００４５】第１３の発明においては、１つのサンプルに複数の特性値がある場合に、テコ比の対角要素を計算し、計算された値を基に検出統計量を求め外れ値を求める。

【００４６】第１４の発明においては、入力された値が回帰分析の手法を適用できる場合、回帰分析の残差を計算し、計算された値を基に検出統計量を求め外れ値を求める。

【００４７】第１５の発明においては、入力された値が正準相関分析モデルのデータの場合、正準相関分析を行い合成変量関数を２個求め、これより合成変量関数値を求め、合成変量関数値からテコ比を計算しテコ比の計算された値を基に検出統計量を求め外れ値を求める。

【００４８】第１６の発明においては、入力された値が複数のグループに分類されていて、複数の要因により判別分析を行う場合に、判別関数値を計算し計算された値を基に検出統計量を求め外れ値を求める。

【００４９】第１７の発明におけるデータ処理装置は、計測手段によりＮ個の値を計測し、この計測された値から、上記外れ値検出方法を実行する外れ値検出手段により、外れ値を検出し、出力手段により外れ値を知らせる。

【００５０】第１８の発明におけるデータ処理装置は、計測手段によりＮ個の値を計測し、この計測された値から、上記外れ値検出方法を実行する外れ値検出手段により、外れ値を検出し、データ処理手段により検出された外れ値を除いた残りの値を用いて所定の処理を実行する。

【００５１】

【実施例】

実施例１．従来例で説明した図１３を再びこの実施例の

9

装置を説明するための図として説明する。図１３で、１は情報処理装置、２はコンピュータ（ＦＤＤ付）、３はディスプレイ・ユニット、４はプリンタ、５はキーボード、６はフロッピーディスクである。この発明のハードウェア構成は従来例と変わらず、プログラム・ルーチンが記憶されたフロッピーディスク６をコンピュータ（ＦＤＤ付）２に挿入し、オペレーション・ソフトを駆動して、情報処理装置１をスタートさせる。プログラム・ルーチンがロードされ、入力待状態となる。キーボード５からデータをキー入力すれば、プログラム・ルーチンが動作し、ディスプレイ３に処理結果を表示し、また、プリンタ４に処理結果をプリントすることになる。

10

【００５２】この実施例では、検出統計量を算出するために数２を使う。

【００５３】

【数２】

$$Ut = n \log \sigma + 2S$$

ただし、n はサンプル数

S は外れ値の候補の個数

σ は $x_1, x_2, \cdots x_n$ をサンプルデータとすると

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

【００５４】この統計量の値が最小になるサンプルの組み合せを見つければよい。図１は本発明の説明のためのフローチャートである。ステップ２０は、キーボード５からのデータを逐次的に入力する入力工程である。例えば、x_1、x_2、x_3、x_4、x_5 の５つのデータを入力する。この場合は、入力するデータの個数をＮとすると、Ｎ＝５となる。ステップ２１は、入力されたデータの大小を比較し、例えば昇順に $x_1 < x_2 < x_3 < x_4 < x_n$ のように並べる。この工程は大小判定工程である。

【００５５】このようにデータを昇順に並べかえることによって、外れ値の候補を見つけることが容易となる。外れ値の候補の個数をｓ（ｓ≧１）とすると、外れ値の候補は、その性質からいって一番大きい値からｓ個、一番小さい値からｓ個、または大きい値と小さい値の両方あわせてｓ個と考えられる。

【００５６】ステップ２２はこれらのデータ群から、本実施例での計算式により検出統計量Ｕｔを計算する算出工程である。外れ値の候補の個数ｓ＝１の場合は、まず、（x_1、x_2、x_3、x_4、x_5）から x_1 を除いた時の検出統計量を計算する。これを検出統計量 $Ut_{(-1)}$ とする。以下同様に x_5 を除いた時を $Ut_{(-5)}$ とする。外れ値の候補の個数ｓ＝２の場合は、x_1 と x_2 を除いた時を $Ut_{(-1,-2)}$ とし、x_4 と x_5 を除いた時を $Ut_{(-4,-5)}$ とし、x_1 と x_5 を除いた時を $Ut_{(-1,-5)}$ とし、小さい値または大きい値から順にサンプルを除いて検出統計量を計算する。このように考えられる組み合せの検出統計量をそれぞれ計算する。ここで外れ値の候補の個数ｓは、予めシステムにより定められているものとする。あるいは、外れ値の候補の個数ｓは、オペレータ、あるいは、プログラムにより指定されるものとする。あるいは、外れ値の候補の個数は計算の度に自由に設定することが可能なものであるとする。

【００５７】尚、ここで与えられる外れ値の候補の個数は、外れ値として必ず見つけなければいけない個数ではない。ここで言う外れ値の候補の個数とは、外れ値として検出する最大の個数を言う。例えば、外れ値の候補の個数ｓ＝２の場合は、外れ値を最大２個見つける場合を言い、外れ値の個数を必ず２個見つけるという意味ではない。従って、外れ値の候補の個数ｓ＝２の場合は、外れ値が０個の場合、外れ値が１個の場合、あるいは、外れ値２個の場合というような結果が考えられる。以下同様に外れ値の候補の個数ｓという場合は、外れ値として検出できる数の最大値を示すものとする。このように、この実施例及び後述する実施例においては、外れ値の数を特定の数に設定する必要はなく、外れ値の数の最大値を指定しておけばよい。

【００５８】ステップ２３は、検出統計量Ｕｔの最小値（Ｕｔｍｉｎ）を見つける段階である。ステップ２４は、外れ値の候補を除かない時の検出統計量 $Ut_{(0)}$ とＵｔｍｉｎを比較する段階である。ステップ２５は、Ｕｔｍｉｎの方が小さい場合、Ｕｔｍｉｎを求めた時のデータの組み合せに含まれていなかった値を外れ値とする。ステップ２６は、$Ut_{(0)}$ が最小となる場合で、この時、外れ値は「ない」とする。ステップ２３からステップ２６までが外れ値検出工程である。ステップ２７は、表示等をする。

【００５９】次に、データを使ってこのフローチャートの流れを説明する。ステップ２０で次の５つのデータを入力する。

５．７１、６．５７、７．２９、８．０６、１３．３２

ステップ２１で入力されたデータを昇順に並べかえる。ステップ２２で数２を使い、Ｕｔの値を計算する。例えば、一番小さな値５．７１を除いた時は、サンプル数ｎ

11

＝４、外れ値の候補の個数ｓ＝１であるので、検出統計量をＵｔ(-1)と表すと、Ｕｔ(-1)＝５．９０８となる。一番大きな値１３．３２を除いた時の検出統計量は、サンプル数ｎ＝４、外れ値の候補の個数ｓ＝１であるので、Ｕｔ(-5)と表すと、Ｕｔ(-5)＝１．４４０となる。５．７１と１３．３２をともに除いた時は、サンプル数ｎ＝３、外れ値の候補の個数ｓ＝２であるので、Ｕｔ＊(-1,-5)と表すと、Ｕｔ(-1,-5)＝２．５０９となる。また、Ｕｔ(-1,-2)、Ｕｔ(-4,-5)、Ｕｔ(-1,-2,-5)、Ｕｔ(-1,-4,-5)を計算すると（即ち、外れ値の候補の個数ｓ＝３の検出統計量Ｕｔ(0)を計算すると）表２のようになる。

12

【００６０】

【表２】

サンプルの組み合わせと検出統計量の値

外れ値の候補 大きい / 小さい	なし	１３．３２	１３．３２ ８．０６
なし	４．９３０ （Ｓ＝０）	（Ｓ＝１） １．４４０ （最小値）	２．６８９ （Ｓ＝２）
５．７１	５．７１ （Ｓ＝１）	２．５０９ （Ｓ＝２）	３．９５７
５．７１ ６．５７	６．９５７ （Ｓ＝２）	４．０９１	

【００６１】ステップ２３で以上で求めたＵｔの最小値Ｕｔｍｉｎを求めると、Ｕｔｍｉｎ＝Ｕｔ(-5)＝１．４４０である。ステップ２４で外れ値の候補を除かない時の検出統計量Ｕｔ(0)＝４．９３０を計算し、ＵｔｍｉｎとＵｔ(0)を比較する。すると、Ｕｔ(-5)＞Ｕｔ(0)は成立しないので、ステップ２５へ行き、Ｕｔ(-5)の時のデータの組み合せを外れ値とする。即ち、１３．３２が外れ値とわかる。ステップ２７で、外れ値１３．３２の表示等出力を行う。

【００６２】このように外れ値の候補の個数が３の場合であっても、検出された外れ値の個数は１つであり、外れ値の候補の個数以内の範囲で外れ値を検出することができる。

【００６３】この例で示した５．７１、６．５７、７．２９、８．０６、１３．３２のデータの場合、１３．３２を外れ値とするのは、竹内　啓（１９８０）「現象と行動の中の統計数理」新曜社でも同様の結果となっている。

【００６４】この実施例の数２に示した検出統計量の式

Ｕｔ＝ｎｌｏｇσ＋２ｓ

は、ＡＩＣ（AKAIKE'S　information　criterion）のアナロジーから考えられた。ｎはサンプル数、ｓは外れ値の候補の個数、σ^2は分散、σは標準偏差である。この式の第１項は、外れ値の候補であるサンプルが除かれると小さくなる傾向がある。というのは、分散σ^2は外れ値を除くと小さくなるからである。また、サンプル数ｎも外れ値の候補として除く数が増えると、ｎ＝（データ数Ｎ）－（外れ値の候補の個数ｓ）であるから、例えば、ｎは５から４、５から３というように小さな値になるからである。第２項は、サンプル数が多くなると増加する。従って、外れ値を除いた時、第１項と第２項の和Ｕｔは最小になると考えることができる。

【００６５】次に、この関係をＧｒｕｂｂｓのデータ１を使って述べる。Ｇｒｕｂｂｓのデータ１を次に示す。

２．０２、２．２２、３．０４、３．２３、３．５９、３．７３、３．９４、４．０５、４．１１、４．１３

総データ数は１０個である。このデータからｌｏｇσを計算した値を表３に、ｎｌｏｇσを計算した値を表４に示し、Ｕｔ＝ｎｌｏｇσ＋２ｓを計算した値を表５に示す。

【００６６】

【表３】

13　　14

Ｇｒｕｂｂｓのデータ１によるｌｏｇσの値

外れ値の候補 大／小	なし	4．13	4．13 4．11	4．13 4．11 4．05
なし	－0．313	－0．318	－0．337	－0．376
2．02	－0．514	－0．516	－0．533	——
2．02 2．22	－0．95	－0．967	——	——
2．02 2．22 3．04	－1．183	——	——	——

【００６７】

✲　✲【表４】

Ｇｒｕｂｂｓのデータ１によるｎｌｏｇσの値

外れ値の候補 大／小	なし	4．13	4．13 4．11	4．13 4．11 4．05
なし	－3．127	－2．858	－2．697	－2．629
2．02	－4．630	－4．127	－3．732	——
2．02 2．22	－7．6	－6．771	——	——
2．02 2．22 3．04	－8．284	——	——	——

【００６８】

※　※【表５】

Ｇｒｕｂｂｓのデータ１によるＵｔ＝ｌｏｇσ+2Ŝの値

外れ値の候補 大きい／小さい	なし	4．13	4．13 4．11	4．13 4．11 4．05
なし	(S＝0) －3．127	(S＝1) －0．858	(S＝2) 1．303	(S＝3) 3．371
2．02	(S＝1) －2．630	(S＝2) －0．127	(S＝3) 2．268	——
2．02 2．22	(S＝2) －3．600 最小	(S＝3) －0．771	——	——
2．02 2．22 3．04	(S＝3) －2．284	——	——	

【００６９】これらをグラフにしたものが図２である。図２のｘ軸は、外れ値の候補の個数ｓであり、ｙ軸はＵｔの値である。ｓに対応するＵｔの値が複数ある場合は、その中の最小のものをプロットした。例えば、表３において外れ値の候補の個数ｓ＝１の場合、Ｕｔは－０．３１８と－０．５１４となるが、－０．５１４を用いてプロットした。図２の一点鎖線で示したグラフ（１）は、Ｕｔ＝ｌｏｇσとした場合を示している。点線で示したグラフ（２）は、Ｕｔ＝ｎｌｏｇσとした場合を示している。実線で示したグラフ（３）は、Ｕｔ＝２ｓとした場合である。太線で表したグラフ（４）は、ｎｌｏｇσと２ｓを加算したＵｔ＝ｎｌｏｇσ＋２ｓの値である。

【００７０】前述したように外れ値の候補の個数ｓが増加するに従って、第１項のｎｌｏｇσは減少することがグラフ（２）よりわかる。そして、第２項の２ｓは増加することが、グラフ（３）よりわかる。グラフ（４）に示すｎｌｏｇσ＋２ｓの値は、ｓ＝２で最小値を取った後増加している。Ｕｔ＝ｎｌｏｇσ＋２ｓの場合、最小値は１ヶである。表５からわかるように、グラフ（４）が最小となるのは、外れ値の候補の個数ｓ＝２であって、その外れ値として２．０２と２．２２を仮定した場合である。これは、Ｋｉｔａｇａｗａの方法でも同じ結果を得ている（ＧｅｎｓｈｉｒｏＫｉｔａｇａｗａ（１

15

979）："On the Use of Aic for the Detection of Outliers", Technometrics, Vol. 21, No. 2）。

【0071】次に、Ut＝nlogσ＋2sとsの関係をもう1つ別の例で述べる。データとして、

−1.40、−0.44、−0.30、−0.24、−＊

16

＊0.22、−0.15、−0.13、0.06、0.10、0.18、0.20、0.48、0.63、1.01

とする。これにより得られたUt＝nlogσ＋2sの値を表6に示す。

【0072】

【表6】

外れ値の候補 大きい／小さい	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	−9.42	−8.32	−6.24	−3.89
−1.40	-11.15	−11.12	−7.65	——
−1.40 −0.44	−8.82	−8.81	——	——
−1.40 −0.44 −0.30	−6.14	——	——	——

【0073】表6を基に、グラフを書くと図3のようになる。図3のx軸は外れ値の候補の個数s、y軸はUt＝nlogσ＋2sの値である。図3からわかるように、この場合もUtの値が最小になった後、sが大きくなるにつれ、Utの値も大きくなっている。また、外れ値の性格より、外れ値の数は総データ数に比して小さな数であると考えられる。よって、以後検出統計量の計算結果を表に示す場合、最小の前後のデータのみを示すことにする。この外れ値の性格を用いることにより、外れ値の候補の数を予め指定することなく、外れ値を検出することも可能である。前述したように、外れ値の数が大きくなるにつれて検出統計量の値も大きくなる。従って、外れ値の候補の数を指定しない場合には、外れ値の候補の数が少ない順に検出統計量を算出し、順に外れ値の候補の数を増やして検出統計量を算出し、その計算した検出統計量が次第に大きくなる場合には、その計算を終了させる。このようにして、外れ値の候補の数が予め指定されていない場合であっても、外れ値を検出することが可能になる。従って、前述したように外れ値の候補の数を予め指定する場合以外に、外れ値の候補の数をシステムやプログラムにより指定せずに、検出統計量の計算結果を比較していくことにより、その計算結果が次第に大きくなることが判明した時点で検出統計量の計算を停止させることにより、外れ値を検出することが可能になる。次に、検出統計量Utが有効であるかどうか検証するために、従来の計算方法による結果と比較したものを実施例2から実施例6で述べる。

【0074】実施例2．この実施例では、Grubbsのデータ2を用い、検出統計量の式としてUtを用いた場合の外れ値について述べる。Grubbsのデータは、全て次の文献より引用している。

"Procedures for detedting outying Observations in samples", Technometrics, Vol. 11, 1−21

Grubbsのデータ2は次の値である（データ数は12）。

0.745、1.832、1.856、1.884、1.914、1.916、1.947、1.949、2.013、2.023、2.045、2.327

原典では、3回の観測値とその平均値が載っているが、ここでは平均値のみを昇順に載せる。検出統計量の計算結果を表7に示す。

【0075】

【表7】

Grubbsのデータ2と検出統計量

外れ値として用いたサンプル	Ut=nlogσ+2s
なし	-2.214
2.327	-9.578
2.045 2.327	-8.244
0.745	-20.497(2)
0.745 1.832	-16.610
0.745 2.327	-22.861(1)
0.745 1.832 2.327	-19.107
0.745 2.045 2.327	-19.145

【0076】表7よりUtが最小値をとるのは、0.745と2.327を外れ値とした場合である。前述したKitagawaの方法も同じ結果となっている。

【0077】実施例3．この実施例は、Grubbsのデータ3を用いた場合について述べる。Grubbsのデータ3は次の値である（データ数は10）。

568、570、570、570、572、572、572、578、584、596

検出統計量の計算結果を表８に示す。

【００７８】

【表８】

Ｇｒｕｂｂｓのデータ３（総データ数１０）と検出統計量

除去候補 / 外れ値として除去されたサンプル	nlogσ+2s
なし	21.109
596	15.975(2)
584 596	12.191(1)
568	21.075
568 570	21.158
568 596	16.319

【００７９】表８よりＵｔが最小値をとるのは、５８４、５９６を外れ値とした場合である。Ｇｒｕｂｂｓによると５９６を外れ値としている。Ｄａｌｌａｓ Ｅ. Ｊｏｈｎｓｏｎ他は、５８４、５９６を外れ値とした。これは、数２で求めた場合と同じである。なお、以後Ｄａｌｌａｓ Ｅ. Ｊｏｈｎｓｏｎ他という場合は、次の資料に基づくものとする。

Ｄａｌｌａｓ Ｅ. Ｊｏｈｎｓｏｎ, Ｓｔｅｐｈｅｎ Ａ. ＭｃＧｕｉｒｅ, ａｎｄ Ｇｅｒｏｇｅ Ａ. Ｍｉｌｌｉｋｅｎ（１９７８）：“Ｅｓｔｉｍａｔｉｎｇ σ^2 ｉｎ ｔｈｅ Ｐｒｅｓｅｎｃｅ ｏｆ Ｏｕｔｏｌｉｅｒｓ”, Ｔｅｃｈｎｏｍｅｔｒｉｃｓ, Ｖｏｌ. ２０, Ｎｏ. ４

【００８０】実施例４．更に、実施例３のデータで、同じデータを重複させてサイズを２倍にしたものを用いた場合を次に示す。データ４は次の値である（データ数は２０）。

５６８、５６８、５７０、５７０、５７０、５７０、５７０、５７０、５７２、５７２、５７２、５７２、５７２、５７２、５７８、５７８、５８４、５８４、５９６、５９６

【００８１】この実施例では、データのサイズを２倍にしたので、外れ値の候補の個数を４（ｓ＝４）とする場合について説明する。外れ値の候補の個数が１の場合は、以下のような組み合せに対して検出統計量を算出することになる。即ち、外れ値の候補の個数が１（ｓ＝１）の場合の統計検出量と、外れ値の候補の個数が２（ｓ＝２）の場合の検出統計量と、外れ値の候補の個数３（ｓ＝３）の場合の検出統計量と、外れ値の候補の個数が４（ｓ＝４）の場合の検出統計量を求める必要がある。外れ値の候補の個数ｓに対応する検出統計量は、以下に示すとおりである。

ｓ＝１　$Ut_{(-1)}$

$Ut_{(-20)}$

ｓ＝２　$Ut_{(-1,-2)}$

$Ut_{(-1,-20)}$

$Ut_{(-19,-20)}$

ｓ＝３　$Ut_{(-1,-2,-3)}$

$Ut_{(-1,-2,-20)}$

$Ut_{(-1,-19,-20)}$

$Ut_{(-18,-19,-20)}$

ｓ＝４　$Ut_{(-1,-2,-3,-4)}$

$Ut_{(-1,-2,-3,-20)}$

$Ut_{(-1,-2,-18,-20)}$

$Ut_{(-1,-18,-19,-20)}$

$Ut_{(-17,-18,-19,-20)}$

【００８２】外れ値の候補の個数が４の場合においても図１に示したフローチャート同様の順に外れ値を検出することが可能である。異なる点は、図１におけるステップ２２において前述したようなｓ＝１からｓ＝４までのそれぞれの検出統計量を算出する点である。このようにして、算出された検出統計量Ｕｔの計算結果を表９に示す。

【００８３】

【表９】

Ｇｒｕｂｂｓのデータ４（総データ数２０）と検出統計量

除去候補 / 外れ値として除去されたサンプル	nlogσ+2s
なし	42.218
596	38.732
596 596	31.950
584 596 596	29.538(2)
584 584 596 596	24.382(1)
568	42.207
568 568	42.150
568 568 570	42.253
568 568 570 570	42.317
568 596	38.887
568 596 596	32.330
568 568 596 596	30.055
568 568 596	38.993
568 568 570 596	39.294

【００８４】表９より外れ値は、５８４、５８４、５９６、５９６である。ＤａｌｌａｓＥ. Ｊｏｈｎｓｏｎ他も同様の結論となっている。

【００８５】尚、外れ値の候補の個数は、入力されたデータの数に基づいて常識的な範囲で任意に設定できるものである。例えば、入力されたデータの数が５（Ｎ＝５）である場合に、外れ値の候補の個数は１又は２（ｓ＝１又は２）とするのが常識的な範囲である。また、入力されたデータの数が多くなれば外れ値の候補の個数も多くする分には差し支えない。このように外れ値の候補の個数は、入力されたデータの数、あるいは、そのシステムにおいて、どの位の精度を要求しているかというシステムの要求に応じて判断されるべきものである。前述した実施例、あるいは、後述する実施例においては、外

19

れ値の数を何個と推定するかという判断は、予めシステムにより定められているか、あるいは、オペレータやプログラムにより任意に指定できるものとする。

【００８６】実施例５．次に、Ｒｏｓｎｅｒのデータを用いた例について述べる。この例は、サイズが５４と比較的大きく、外れ値も多く存在すると考えられるケースである。次に、データを示す。

－０．２５、０．６８、０．９４、１．１５、１．２０、１．２６、１．２６、１．３４、１．３８、１．４３、１．４９、１．４９、１．５５、１．５６、１．５８、１．６５、１．６９、１．７０、１．７６、１．７７、１．８１、１．９１、１．９４、１．９６、１．９９、２．０６、２．０９、２．１０、２．１４、２．１５、２．２３、２．２４、２．２６、２．３５、２．３＊

20

＊７、２．４０、２．４７、２．５４、２．６２、２．６４、２．９０、２．９２、２．９２、２．９３、３．２１、３．２６、３．３０、３．５９、３．６８、４．３０、４．６４、５．３４、５．４２、６．０１

この、Ｒｏｓｎｅｒのデータは次の文献からとった。

Ｂｅｒｎａｒｄ Ｒｏｓｎｅｒ（１９７７）：“Ｐｅｒｃｅｎｔａｇｅ Ｐｏｉｎｔ ｆｏｒ ａ Ｇｅｎｅｒａｌｉｚｅｄ ＥＳＤ Ｍａｎｙ－Ｏｕｔｌｌｅｒ Ｐｒｏｃｅｄｕｒｅ”，Ｔｅｃｈｎｏｍｅｔｒｉｃｓ，Ｖｏｌ．２５，Ｎｏ．２

次に、Ｕｔの計算結果を表１０に示す。

【００８７】

【表１０】

Ｒｏｓｎｅｒのデータと検出統計量

外れ値として疑いたサンプル ＼ 検出統計量	nlogσ+2s
なし	8.564
-0.25	8.379
-0.25 0.68	9.607
6.01	5.415
5.42 6.01	3.008
5.34 5.42 6.01	-0.234
-0.25 6.01	4.939
-0.25 5.42 6.01	2.225
-0.25 0.68 5.42 6.01	3.346
4.64 5.34 5.42 6.01	-2.009
4.30 4.64 5.34 5.42 6.01	-3.305(1)
3.68 4.30 4.64 5.34 5.42 6.01	-2.995(2)
3.59 3.68 4.30 4.64 5.34 5.42 6.01	-2.681
3.30 3.59 3.68 4.30 4.64 5.34 5.42 6.01	-1.720
-0.25 5.34 5.42 6.01	-1.518

【００８８】Ｒｏｓｎｅｒは、外れ値が最大１０あると仮定して検定した。危険率５％で、５．３４、５．４２、６．０１を外れ値とした。表１０は、この実施例による検出統計量の計算結果を示す表である。前述したように検出統計量の計算結果を表に示す場合には、最小の値の前後のデータのみを示してある。この表１０からわかるように、計算式Ｕｔを用いた場合、外れ値は４．３０、４．６５、５．３４、５．４２、６．０１である。この場合には、外れ値として５つの外れ値が検出されているが、Ｒｏｓｎｅｒが仮定したように外れ値が最大１０個あると仮定した場合であっても、あるいは、外れ値の候補の数を指定せずに外れ値の候補の数を増やす毎に計算された検出統計量を比較することにより自動的に外れ値を検出した場合のいずれの場合においても、結果はこの５つの外れ値を検出する。この実施例においては、５つの外れ値を検出したが、もしこの方法で外れ値が３個までとすると、Ｒｏｓｎｅｒと一致している。

【００８９】実施例６．正規乱数、指数乱数、一様乱数をサンプルデータとした場合について述べる。正規乱数、指数乱数は、外れ値が現れる可能性があるが、一様乱数からは外れ値は出て欲しくない。正規乱数データは（ｎ＝１０，～Ｎ（０，１））より、

－２．６６６、－１．２７２、－０．０４２、０．１４０、０．２７３、０．４１５、０．４６７、１．１６０、１．６７２、１．６７３

である。Ｕｔの計算結果を表１１に示す。

【００９０】

【表１１】

21

正規乱数と検出統計量データ(n=10, ～N(0,1)))

検出統計量 / 外れ値として除いたサンプル	nlogσ+2s
なし	2.294
1.167	3.475
1.672 1.673	4.998
-2.666	0.744(2)
2.666 -1.272	0.436(1)
-2.666 1.673	2.320
1.160 1.672 1.673	6.546
-2.666 1.673	3.329
-2.666 -1.272 1.673	2.017
-2.666 -1.272 -0.042	2.559

【００９１】表１１より外れ値は－２．６６６、－１．２７２である。

【００９２】次に、指数乱数を用いた場合について示す。データは、竹内「現象と行動の中の統計数理」（新曜社）からとった。

０．００３、０．０２１、０．１６１、０．１７８、０．１８０、０．２１０、０．２４９、０．４１３、０．４９４、０．５６２、０．６１３、０．８７９、０．９８１、１．０５９、１．１３１、１．２６４、２．３６７、３．６６９、３．８２６、４．１９３

総データ数は２０である。Ｕｔの計算結果を表１２に示す。

【００９３】

【表１２】

指数乱数と検出統計量

検出統計量 / 外れ値として除いたサンプル	nlogσ+2s
なし	5.041
4.193	3.894
4.193 3.826	2.074
4.193 3.826 3.669	-3.091(3)
4.193 3.826 3.669 2.367	-6.402(1)
0.003	6.889
0.003 0.021	8.706
0.003 0.021 0.161	10.573
0.003 0.021 0.161 0.178	12.405
4.193 0.003	5.896
0.003 4.193 3.826	4.261
0.003 4.193 3.826 3.669	-0.674
0.003 0.021 4.193	7.866
0.003 0.021 4.193 3.826	6.415

22

【００９４】表１２より外れ値は、４．１９３、３．８２６、３．６６９、２．３６７である。

【００９５】次に、［０、１］の一様乱数を用いた場合について述べる。データ数１０でデータは次の通りである。

０．２８３、０．４７０、０．６４３、０．６８８、０．９１６、０．９３０、０．９４５、０．９５３、０．９７３、０．９９５

Ｕｔの計算結果を表１３に示す。

【００９６】

【表１３】

［０，１］の一様乱数と検出統計量

検出統計量 / 外れ値として除いたサンプル	nlogσ+2s
なし	-14.484(1)
0.995	-11.003
0.995 0.973	-7.536
0.283	-13.649(2)
0.283 0.470	-12.533
0.283 0.995	-9.878
0.283 0.470	-4.109
0.283 0.995 0.973	-6.103
0.283 0.470 0.995	-8.435
0.283 0.470 0.643	-10.439

【００９７】表１３より外れ値の候補がない場合が最小となっているので、一様乱数の場合、外れ値の候補はない。一様乱数については、更に１ケース試みたが同様に外れ値の候補はなかった。一様乱数という性格上、外れ値の候補なしということは望ましい結果である。

【００９８】実施例７．従来の技術に出ているデータについて、Ｕｔを用いて検出統計量を求める。データは、

５．７１、６．５７、７．２９、８．０６、１０．００、１５．００

である。結果は表１４のようになる。

【００９９】

【表１４】

サンプルの組み合せと検出統計量の値

ケース	n	s	検出統計量
除かないとき	6	0	6.77　$Ut_{(0)}$
5.71を除いたとき	5	1	7.55　$Ut_{(-1)}$
5.71と15.00ともに除いたとき	4	2	4.99　$Ut_{(-1,-6)}$
15.00を除いたとき	5	1	3.90　$Ut_{(-6)}$(採択)

【０１００】表１４よりＵｔは１５．００を除いた時、最小値となることがわかり、１５．００を外れ値とする。

【０１０１】実施例８．従来技術で述べたマスク効果のデータについて、数２に示した計算式を用いて検出統計量を求める。データは、

＊5．71、6．57、7．29、8．06、14．80、15．00

である。結果は表１５のようになる。

【０１０２】

【表１５】

サンプルの組み合せと検出統計量の値

ケース	n	s	検出統計量
除かないとき	6	0	8.06　$Ut_{(0)}$
5.71を除いたとき	5	1	8.61　$Ut_{(-1)}$
5.71，15.00を除いたとき	4	2	3.76　$Ut_{(-1,-6)}$
15.00を除いたとき	5	1	7.90　$Ut_{(-6)}$
15.00，14.80を除いたとき	4	2	3.44　$Ut_{(-6,-5)}$
5.71，6.57を除いたとき	4	2	9.15　$Ut_{(-1,-2)}$
5.71，15.00，14.80を除いたとき	3	3	4.51　$Ut_{(-1,-6,-5)}$
5.71，6.57，15.00を除いたとき	3	3	9.65　$Ut_{(-1,-2,-6)}$

【０１０３】表１５より$Ut_{(-6,-5)}$の時、最小値となることがわかり、１５．００、１４．８０を外れ値とする。このように数２に示した計算式を用いれば、マスク効果を回避することができる。

【０１０４】実施例９．図４は、この実施例を説明するための図である。この実施例のデータ処理装置は、センサー等の計測手段を有し、これより得られた測定値から、外れ値を検出する外れ値検出手段を有する。次に、外れ値がある場合は、これを除いたデータで平均値を求めるデータ処理手段を有す。次に、実際の適用例について述べる。センサーから得られる測定値を、一定時間間隔ごとに５個測定し外れ値を検出し、これを除いた平均値を測定値とすることを考える。x_1、x_2、x_3、x_4、x_5が、測定データとして得られる。検出統計量Ｕｔを計算し、外れ値を求める。外れ値があれば外れ値を除き偏りのない平均値を求めることができる。データ

は時刻ｔ１、ｔ２、ｔ３、ｔ４について表１６に示す。

【０１０５】

※【表１６】

時刻	データ					平均値
t1	3.23,	1.24,	2.03,	2.86,	1.02	2.08
t2	2.06,	0.74,	2.19,	2.39,	2.73	2.01
t3	2.90,	0.08,	2.55,	2.23,	2.53	2.06
t4	1.83,	2.36,	3.36,	2.44,	1.95	2.39

【０１０６】時刻ｔ１のデータについて検出統計量Ｕｔを求める。除かない時、$Ut_{(0)}$＝－０．７１１となる。１．０２を除いた時、$Ut_{(-1)}$＝０．９５２、３．２３を除いた時、$Ut_{(-6)}$＝０．７０９となる。１．０２、３．２３ともに除いた時、$Ut_{(-1,-6)}$＝２．７６０となる。これを表１７にまとめると次のようになる。$Ut_{(-1,-2)}$、$Ut_{(-4,-6)}$も考えられるがこの実施例では影響がないので表に示すのを省略する。以下の表でも同様に影響がないものは表示しないことにする。

【０１０７】

【表１７】

外れ値の候補 ＼ 大きい方 / 小さい方	なし	3.23
なし	－0.711	0.709
1.02	0.952	2.760

【０１０８】外れ値がない時のＵｔが－０．７１１と最小である。従って外れ値はない。時刻ｔ２のデータについてＵｔを求め、表１８に示す。

【０１０９】

【表１８】

外れ値の候補 ＼ 大きい方 / 小さい方	なし	2.73
なし	－1.969	0.205
0.74	－3.524	－2.504

※【０１１０】０．７４を除いた時のＵｔが－３．５２４と最小である。従って０．７４を外れ値とする。時刻ｔ３のデータについてＵｔを求め、表１９に示す。

【０１１１】

【表１９】

外れ値の候補 ＼ 大きい方 / 小さい方	なし	2.90
なし	0.057	2.112
0.08	－3.753	－1.765

【０１１２】従って、０．０８を外れ値とする。次項ｔ４のデータについてＵｔを求め、表２０に示す。

【０１１３】

【表２０】

外れ値の候補 ＼ 大きい方 / 小さい方	なし	3.36
なし	－3.092	－3.388
1.36	－0.652	－0.616

【０１１４】従って、３．３６を外れ値とする。更新されたデータ及び平均値は、表２１のようになる。

【０１１５】

【表２１】

時刻	更新されたデータ					更新された平均値
t1	3.23,	1.24,	2.03,	2.86,	1.02	2.08
t2	2.06,	2.19,	2.36,	2.73		2.33
t3	2.90,	2.55,	2.23,	2.53		2.55
t4	1.83,	2.36,	2.44,	1.95		2.15

【０１１６】実施例１０．この実施例は、外れ値検出手段を有し、外れ値を知らせる出力手段を有するデータ処理装置について述べる。入力工程から表２２のようなデータが得られ、このデータからＵｔを計算することにより外れ値を検出する。表２２のデータを図示すると図５のようになる。

【０１１７】

【表２２】

27 28

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y	1	1.5	−1	−2	−1.5	0	−3	0	1.5	1.4	0	−1	0	

【０１１８】このデータについて検出統計量Ｕｔを求めると表２３のようになる。

＊【０１１９】

＊【表２３】

外れ値の大きい方 / 採択 / 小さい方	なし	4	4，1.5
なし	7.30	5.65	6.89
−3	7.39	5.23	6.54
−3，−2	8.22	5.77	6.54

【０１２０】表２３より４、－３を外れ値とする。従来は、入力工程から得られたデータをディスプレー等に図５に示すようなグラフを表示し、人が目視によって外れ値と思われる値をピック・アップしてから計算し、確かめていた。本発明による装置により外れ値を自動的に検出し、工程環境・条件に異常があったかどうかの確認を行うことができる。

【０１２１】実施例１１．図６はこの実施例を説明するための図である。図６で示した外れ値検出方法は、図１の入力工程と大小判定工程の間に、入力したデータを加工する加工工程が追加されたものである。この実施例では、時間とともに増加する傾向がある特性値を出力する装置からデータをうけとり、加工工程により時間とともに増加する傾向を除いたデータから、外れ値を検出する装置について述べる。データを表２４に示す。これを図示すると図７の様になり、データは時間とともに増加する傾向があることが分かる。

【０１２２】

【表２４】

t	1	2	3	4	5	6	7	8	9	10
y	2	4	3	6	6	5	6	11	6	9

【０１２３】そこで、最小２乗法により傾向直線を求めると次のようになる。

$y = 1.93 + 0.70t$

データから傾向直線の値を引くことにより、下のようなデータとなる。

－０．６３、０．６７、－１．０３、１．２７、０．５７、－１．１３、－０．８３、３．４７、－２．２３、０．０７

これを図示すると、図８のようになる。この補正されたデータから、Ｕｔを計算すると表２５のようになる。

【０１２４】

【表２５】

外れ値の採択 / 大きい方 / 小さい方	なし	3.47
なし	4.13	2.29
−2.23	4.92	2.69

【０１２５】従って、補正されたデータ３．４７に対応するｔ＝８のｙ＝１１を外れ値とする。この様に、時間とともに増加する傾向のあるデータから、増加する傾向を補正することにより、実施例１の外れ値検出方法を適用することができる。

【０１２６】実施例１２．この実施例は、一つのサンプルに複数の特性値がある場合に、加工工程においてテコ比（$X(X^T X)^{-1} X^T$）の対角要素を計算し、これをもとに外れ値を検出する装置について説明する。データおよび計算されたテコ比を表２６に示す。

【０１２７】

【表２６】

29

no.	x1	x2	x3	x4	テコ比
1	7	26	6	60	0.48
2	1	29	15	52	0.28
3	11	56	8	20	0.13
4	11	31	8	47	0.24
5	7	52	6	33	0.34
6	11	55	9	22	0.12
7	3	71	17	6	0.36
8	1	31	22	44	0.38
9	2	54	18	22	0.19
10	21	47	4	26	0.67
11	1	40	23	34	0.37
12	11	66	9	12	0.19
13	10	68	8	12	0.24
14	30	85	30	79	0.99

30

＊【０１２８】テコ比を用いて検出統計量Ｕｔを求めると表２７のようになる。

【０１２９】

【表２７】

10

＊

外れ値の候補 ＼ 大きい方 / 小さい方	なし	0.99	0.99 0.67
なし	−20.75	−22.99	−22.85
0.12	−17.37	−19.50	−19.39
0.12, 0.13	−14.02	−16.08	−16.12

【０１３０】－２２．９９が最小値である。従って０．９９つまりサンプルｎｏ．１４を外れ値とする。テコ比が大きいデータは、全体に与える影響が大きいので、外れ値か否か容易に判定できる外れ値検出装置があることは有効である。

【０１３１】実施例１３．この実施例は、入力されたデ※

※ータが回帰分析のモデルの場合、加工工程において回帰分析の残差を求め、これをもとに外れ値を検出する装置について説明する。データおよび回帰式により求めた残差は表２８のようになる。

【０１３２】

【表２８】

no.	x1	x2	x3	y	残差
1	7	26	50	78.5	0.346
2	1	29	52	74.3	1.545
3	11	56	20	104.3	−1.874
4	11	31	47	87.6	−1.783
5	7	52	33	95.9	−0.322
6	11	55	22	109.2	3.948
7	3	71	6	102.7	−1 339
8	1	31	44	72.5	−3.186
9	2	54	22	93.1	1.288
10	21	47	26	115.9	0.246
11	1	40	34	83.8	1.993
12	11	66	12	113.3	1.171
13	10	68	12	109.4	−2.033

【０１３３】残差を用いて検出統計量Ｕｔを求めると表２９のようになる。

★【０１３４】

★【表２９】

外れ値の候補 ＼ 大きい方 / 小さい方	なし	3.948	3.948 1.993
なし	8.459	7.723	8.573
−3.186	8.762	7.873	8.816
−3.186, −2.033	9.746	8.939	10.030

【０１３５】Ｕｔ７．７２３が最小である。従って残差 50 ３．９４８つまりサンプルｎｏ．６が外れ値となる。

31

【０１３６】実施例１４．この実施例は、入力されたデータの特性値が複数あり正準相関分析モデルを適応できる場合、加工工程において次に示すようにデータを加工し、これを用いて外れ値を検出する装置について説明する。表３０のようなデータについて考える。

【０１３７】

【表３０】

	特性値		要因			
no.	y1	y2	x1	x2	x3	x4
1	78.5	21.0	7	26	6	60
2	74.3	22.0	1	29	15	52
3	104.3	23.0	11	56	8	20
4	87.6	24.0	11	31	8	47
5	95.9	23.1	7	52	6	33
6	109.2	19.6	11	55	9	22
7	102.7	18.7	3	71	17	6
8	72.5	23.3	1	31	22	44
9	93.1	23.8	2	54	18	22
10	115.9	19.4	21	47	4	26
11	83.8	28.8	1	40	23	34
12	113.3	19.8	11	66	9	12
13	109.4	19.1	10	68	8	12

【０１３８】このデータに正準相関分析を行い、ｙ１，ｙ２の合成変量関数が２個求まる。この合成変量関数を用いて合成変量関数値が求まる。合成変量関数値をもとにテコ比を計算する。テコ比は次のようになる。

０．２７、０．３１、０．１８、０．１２、０．１０、０．２０、０．１９、０．３０、０．１２、０．３０、０．６３、０．１３、０．１６

このテコ比について検出統計量Ｕｔを計算すると表３１のようになる。

【０１３９】

【表３１】

外れ値 大きい方 / 小さい方	なし	0.63
なし	−26.00	−29.03
0.10	−22.03	−24.96

【０１４０】従ってテコ比０．６３を外れ値とする。これはｎｏ．１１のサンプルである。

【０１４１】実施例１５．この実施例は、入力されたデータが２グループに特性値が分類されていて、加工工程において複数の要因により判別分析を行い、この加工されたデータをもとに各グループでの外れ値を検出する装置について述べる。データは表３２に示す。

【０１４２】

【表３２】

32

データ no.		x1	x2	判別関数値
グループ1	1	6	0	−3.80
	2	0	2	−0.56
	3	0	3	−0.84
	4	1	2	−1.20
	5	1	3	−1.48
	6	1	4	−1.76
グループ2	7	4	0	−2.53
	8	4	1	−2.82
	9	5	0	−3.17
	10	5	1	−3.45
	11	5	2	−3.73
	12	0	4	−1.12

【０１４３】データをプロットすると図９のようになる。図よりグループ１では、サンプルｎｏ．１が、グループ２ではサンプルｎｏ．１２が外れ値のようである。データについて判別分析を実施し、判別関数を求めると、

ｙ＝−０．６３４＊ｘ１−０．２８１＊ｘ２

となる。この判別関数を用いて判別関数値を計算すると例えば、ｎｏ．１の場合、

判別関数値＝−０．６３４＊６−０．２８１＊０＝−３．８０

となる。判別関数値を表３２の右欄に載せた。グループ１の判別関数値についてＵｔを求めると表３３のようになる。

【０１４４】

【表３３】

外れ値 大きい方 / 小さい方	なし	−0.56
なし	0.33	2.19
−3.80	−2.24	−0.33

【０１４５】従って判別関数値−３．８０、サンプルｎｏ．１を外れ値とする。グループ２の判別関数値についてＵｔを求めると表３４のようになる。

【０１４６】

【表３４】

外れ値 大きい方 / 小さい方	なし	−1.124
なし	−1.00	−2.24
−3.73	0.94	−0.24

【０１４７】従って判別関数値−１．１２、サンプルｎｏ．１２を外れ値とする。

【０１４８】実施例１６．この実施例は、入力されたデータが３グループに特性値が分類されていて、加工工程において複数の要因により判別分析を行い、このデータ

33　　34

にもとづいて各グループでの外れ値を検出する装置について述べる。データを表３５に示す。　＊【０１４９】　＊【表３５】

	データ no.	x1	x2	判別関数値 f_1	f_2	ユークリッド距離 d
グループ１	1	0	3	3.45	-1.21	3.65
	2	0	4	4.60	-1.62	4.88
	3	1	2	1.46	-1.33	1.97
	4	1	3	2.61	-1.74	3.14
	5	1	4	3.76	-2.14	4.33
グループ２	6	4	0	-3.39	-2.11	3.99
	7	5	0	-4.23	-2.63	4.98
	8	4	1	-2.24	-2.51	3.36
	9	5	1	-3.08	-3.34	4.33
	10	4	2	-1.08	-2.91	3.10
	11	5	2	-1.93	-3.44	3.94
グループ３	12	4	5	2.37	-4.13	4.76
	13	4	6	3.52	-4.53	5.74
	14	5	5	1.52	-4.55	4.89
	15	5	6	2.67	-5.06	5.72
	16	6	5	0.67	-5.18	5.22
	17	6	6	1.83	-5.58	5.87
	18	0	2	2.30	-0.81	2.44

【０１５０】データについて判別分析を実施し判別関数値を求めると、表３５の右側、判別関数値の欄のようになる。３グループあるので判別関数値は２組得られる。一般に判別関数値は（グループ数－１）組得られる。この２組の判別関数値から次の式に基づいてユークリッド距離が求められ、これを表３５のユークリッド距離の欄に示す。　※

※判別関数値のユークリッド距離ｄ（＝（$f_1^2 + f_2^2$）$^{1/2}$）

判別関数値のユークリッド距離について各グループごとにＵｔを求めると、Ｕｔは次のようになる。グループ１のＵｔを表３６に示す。

【０１５１】　【表３６】

グループ１

外れ値 大きい方 / 小さい方	なし	4.88
なし	0.02	1.41
1.97	0.34	1.84

【０１５２】Ｕｔは、外れ値の候補がない場合が最小である。従って外れ値はない。グループ２のＵｔを表３７に示す。　★

★【０１５３】　【表３７】

グループ２

外れ値 大きい方 / 小さい方	なし	4.98
なし	-2.92	-2.01
3.10	-1.16	-0.22

【０１５４】Ｕｔは、外れ値の候補がない場合が最小である。従って外れ値はない。グループ３のＵｔを表３８に示す。

【０１５５】　【表３８】

グループ３

外れ値 大きい方 / 小さい方	なし	5.87	5.87 5.74
なし	0.67	2.65	4.62
2.44	-3.10	-0.49	2.03
2.44 , 4.76	-0.94	1.88	4.77

【０１５６】従ってユークリッド距離が２．４４、サンプルｎｏ．１８を外れ値とする。なお、この実施例では特性値が３グループの場合について述べたが、３グループ以上についても同様に行える。

【０１５７】実施例１７．この実施例では、上記実施例で示した数２とは異なる計算式で検出統計量を求める場合について述べる。検出統計量Ｕｔａの計算式は数３を用いる。

【０１５８】

【数３】

$$Uta = n \log \alpha^2 + \frac{b_2}{2} + 2S$$

ただし n はサンプル数

S は外れ値の候補の個数

α は $x_1, \cdots, x_n$ をサンプルデータとすると

$$\alpha^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$b_2 = \frac{\sum (x_i - \bar{x})^4}{\alpha^2}$$

＊

除いたサンプル	$n\log\sigma+b_2/2+2s$	参考 $n\log\sigma+2s$
なし	-7.267	-9.419
1.01	-5.860	-8.323
-1.40	-9.770	-11.153 (1)
1.01, -1.40	-10.101 (1)	-11.117
1.01, 0.63	-3.577	-6.237
-1.40, -0.44	-7.421	-8.815
-1.40,-0.44,1.01	-7.825	-8.806
-1.40,1.01,0.63	-8.575	-9.561

【０１５９】データとして（総データ数１５）、

－１．４０、－０．４４、－０．３０、－０．２４、－０．２２、－０．１３、－０．１５、０．０６、０．１０、０．１８、０．２０、０．３９、０．４８、０．６３、１．０１

を用いる。計算結果は表３９のようになる。

【０１６０】

【表３９】

【０１６１】従って、表より１．０１、－１．４０を外れ値とする。表３９の参考の欄を見るとわかるように、Ｕｔ＝ｎｌｏｇσ＋２ｓを用いると－１．４０を外れ値としているので、Ｕｔａでは１個多く外れ値を指定している。ところが、他の多くのデータでは、数２の計算式を用いた場合と同様の結果を得ている。このことより、数３の計算式は、場合によっては１個多く外れ値を検出するという特徴がある。尚、数３の第２項のｂ２／２は、「竹内」の正規分布のあてはまりのよさの補正の指標を参考とした。

竹内（竹内　啓（１９７６）：“情報統計量の分布とモデルの適切さの基準”、「数理科学」、ＮＯ．１５３サイエンス社、１２－１８）によれば正規分布モデルの適切さを表す統計量（以下竹内の統計量）Ｔ，は、z_i（$i=1, \cdots, n$）をサンプル数ｎのデータ、ｚをz_iの平均として次のようになる。

$T_1 = -\log\sigma - b_2/2n$

ここで、

$\sigma^2 = \{\Sigma (z_i - z)^2\}/n$

$b_2 = \{\Sigma (z_i - z)^4\}/n\sigma^4$

37

この竹内の統計量の値が大きいほど適切な正規分布モデルに近い。

【０１６２】正規分布モデルａと正規分モデルｂの２つがあり、正規分布モデルａの分散は正規分布モデルｂの分散よりも小さな値を示す場合、正規分布モデルａの方が正規分布モデルｂよりもデータz_iが平均＝０に近い値を多く示す。上記竹内の統計量を求める式の第２項にあるｂ$_2$／２ｎは、補正項と呼ばれているものであり、第１項にあるｌｏｇσの値を補正する意味を持っているものである。従って、竹内の統計量は第１項にあるｌｏｇσの値が大きく影響するものである。従って、分散σの値によってこの竹内の統計量の特徴付けがなされる。従って、分散が小さいほど竹内の統計量の値が大きくなり、この竹内の統計量の値が大きいほど正規分布モデルｂよりも正規分布モデルａに近いパターン（即ち、分散の小さいパターン）を示すことになる。

【０１６３】実施例１８．以下、この実施例１８から実施例４６までは、検出統計量を求めるための計算式を図１０に示す回帰分析説明変数選択基準を基礎にして作成している。前述した数２及び数３はＡＩＣを基礎にして考えたものである。ＡＩＣは、回帰分析説明変数選択基準の一例である。従って、以下の実施例１８から実施例４６までは、ＡＩＣよる回帰分析説明変数選択基準以外の回帰分析説明変数選択基準を基礎にして、検出統計量を求める場合においても前述した実施例と同様な効果を奏することができる点をについて説明する。実施例１８から実施例４６までに示す検出統計量の計算式数４から＊

38

＊数３２は、図１０に示す回帰分析説明変数選択基準を基礎にして考えられたものであり、これらの計算式は、大きく分けて２つのタイプに分類できる。第１のグループは、前述した実施例までと同じ形式で、

ｎｌｏｇσ＋第２項（＋第３項）

である。第２のグループは、乗算タイプで、

調整因子×σ

である。また、データとして、

－１．４０、－０．４４、－０．３０、－０．２４、－０．２２、－０．１５、－０．１３、０．０６、０．１０、０．１８、０．２０、０．３９、０．４８、０．６３、１．０１

を、この実施例以後の全ての実施例で用いる。これを図示すると図１１のようになる。同じデータに対して、計算式の違いにより求まる外れ値の数が異なっている。よって、外れ値を多く出さなくてもよい適用業務と、外れ値を多く出したい業務により、計算式を選んで使うことができる。

【０１６４】検出統計量Ｓｔの計算式を、数４に示す。

【０１６５】

【数４】

$$St = \frac{(n-2)(n-1)}{(n-S-2)(n-S-1)} \alpha$$

【０１６６】Ｓｔを計算すると表４０のようになる。

【０１６７】

【表４０】

外れ値の 大きい方 候補 / 小さい方	なし	１．０１	１．０１ ０．６３
なし	０．５３４	０．５６５	０．６６７
－１．４０	０．４６２	０．４５８	０．５３７
－１．４０ －０．４４	０．５４７	０．５７２	０．７５２

【０１６８】従って、－１．４０と１．０１を外れ値とする。

【０１６９】実施例１９．検出統計量Ｆｔの計算式を、数５に示す。

【０１７０】

【数５】

$$Ft = \frac{n+S}{n-S} \alpha$$

【０１７１】Ｆｔを計算すると表４１のようになる。

【０１７２】

【表４１】

39　　40

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．83
なし	0．53	0．55	0．62
−1．40	0．45	0．43	0．46
−1．40 −0．44	0．51	0．49	0．54

【０１７３】従って、－１．４０と１．０１を外れ値と　10＊する。

【０１７４】実施例２０．検出統計量Ｔｔの計算式を、数６に示す。

【０１７５】

【数６】

$$Tt = \frac{n^2 - n - S - 2}{n(n-S-2)(n-S-1)}\alpha$$

【０１７６】Ｔｔを計算すると表４２のようになる。

【０１７７】

＊　【表４２】

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．63
なし	0．041	0．046	0．059
−1．40	0．038	0．041	0．052
−1．40 −0．44	0．048	0．055	0．079

【０１７８】従って、－１．４０を外れ値とする。

【０１７９】実施例２１．検出統計量ＴＩｔの計算式を、数７に示す。

【０１８０】

【数７】

※　$$TIt = n \log \alpha + 2S + \frac{S(S-2)}{n}$$

【０１８１】ＴＩｔを計算すると表４３のようになる。

【０１８２】

※30　【表４３】

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．63
なし	−9．42	−8．39	−6．24
−1．40	−11．22	−11．12	−9．31
−1．40 −0．44	−8．82	−8．56	−6．49

【０１８３】従って、－１．４０を外れ値とする。

【０１８４】実施例２２．検出統計量Ｗｔの計算式を、数８示す。

【０１８５】

【数８】

$$Wt = S \log n + n \log \alpha$$

【０１８６】Ｗｔを計算すると表４４のようになる。

【０１８７】

【表４４】

41　　42

外れ値の 大きい方 候補 ＼ 小さい方	なし	1．01	1．01 0．63
なし	－9．42	－7．78	－5．11
－1．40	－10．51	－9．99	－8．11
－1．40 －0．44	－7．69	－7．35	－5．63

【０１８８】従って、－１．４０を外れ値とする。

【０１８９】実施例２３．検出統計量Ｐｔの計算式を、数９示す。

【０１９０】

【数９】

$$Pt = \frac{1}{(n-S-1)^2}\alpha$$

＊【０１９１】Ｐｔを計算すると表４５のようになる。

【０１９２】

【表４５】

＊

外れ値の 大きい方 候補 ＼ 小さい方	なし	1.01	1.01 0.63
なし	0.002722	0.0033	0.0045
－1.40	0.002714	0.0031	0.0043
－1．40 －0．44	0.00373	0.0045	0.0069

【０１９３】従って、－１．４０を外れ値とする。

【０１９４】実施例２４．検出統計量Ｈｔ計算式を、数１０示す。

【０１９５】

【数１０】

※

$$Ht = \frac{(n-1)(n+S+1)}{(n+1)(n-S-2)}\alpha$$

30　【０１９６】Ｈｔを計算すると表４６のようになる。

【０１９７】

※　【表４６】

外れ値の 大きい方 候補 ＼ 小さい方	なし	1．01	1．01 0．63
なし	147.14	135.69	135.89
－1．40	110.85	93.36	89.37
－1．40 －0．44	111.45	85.18	96.28

【０１９８】従って、－１．４０、１．０１、０．６３を外れ値とする。

【０１９９】実施例２５．検出統計量Ｕ．１ｔの計算式を、数１１示す。

【０２００】

【数１１】

$$U.1t = n\log\alpha + 2 * S * 1.1$$

【０２０１】Ｕ．１ｔを計算すると表４７のようになる。

【０２０２】

【表４７】

43　　44

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．63
なし	-9.42	-8.32	-6.24
－1．40	-11.15	-11.12	-9.56
－1．40 －0．44	-8.82	-8.81	-7.22

【０２０３】従って、－１．４０を外れ値とする。

【０２０４】実施例２６．検出統計量U．９ｔの計算式を、数１２示す。

【０２０５】

【数１２】

$$U.9t = n\ log\alpha + 2 * S * 0.9$$

【０２０６】U．９ｔを計算すると表４８のようになる。

【０２０７】

【表４８】

*

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．63
なし	-9.42	-8.52	-6.64
－1．40	-11.35	-11.52	-10.16
－1．40 －0．44	-9.22	-9.41	-8.02

【０２０８】従って、－１．４０と１．０１を外れ値とする。

【０２０９】実施例２７．検出統計量Ｕｕｔの計算式を、数１３示す。

【０２１０】

【数１３】

※

$$Uut = \frac{n-1}{n-S}\alpha$$

【０２１１】Ｕｕｔを計算すると表４９のようになる。

【０２１２】

【表４９】

※

外れ値の 大きい方 候補 / 小さい方	なし	1．01	1．01 0．63
なし	0.498	0.478	0.498
－1．40	0.391	0.341	0.334
－1．40 －0．44	0.407	0.356	0.358

【０２１３】従って、－１．４０、１．０１、０．６３を外れ値とする。

【０２１４】実施例２８．検出統計量Ｂｔの計算式を、数１４示す。

【０２１５】

【数１４】

$$Bt = \frac{n+1}{n-S-1}\alpha$$

【０２１６】Ｂｔを計算すると表５０のようになる。

【０２１７】

【表５０】

45　　46

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.610	0.598	0.637
－1.40	0.489	0.438	0.444
－1.40,-0.44	0.522	0.473	0.501

【０２１８】従って、－１．４０と１．０１を外れ値とする。

*

$$Dt = \frac{(n+1)(n+S+1)}{n-S}\alpha$$

【０２１９】実施例２９．検出統計量Ｄｔの計算式を、数１５示す。

【０２２０】

【数１５】

【０２２１】Ｄｔを計算すると表５１のようになる。

【０２２２】

【表５１】

*

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	8.539	8.243	8.604	9.360
－1.40	6.734	5.911	5.833	6.063
－1.40,-0.44	7.356	6.212	6.304	6.954

【０２２３】従って、－１．４０、１．０１、０．６３を外れ値とする。

※

$$Gt = \frac{n+S+1}{n-S+1}\alpha$$

【０２２４】実施例３０．検出統計量Ｇｔの計算式を、数１６示す。

【０２２５】

【数１６】

【０２２６】Ｇｔを計算すると表５２のようになる。

【０２２７】

【表５２】

※

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.534	0.547	0.607	0.702
－1.40	0.447	0.417	0.438	0.482
－1.40,-0.44	0.498	0.466	0.501	0.580

【０２２８】従って、－１．４０と１．０１を外れ値とする。

$$Qt = \frac{1}{(n-S)^2}\alpha$$

【０２２９】実施例３１．検出統計量Ｑｔの計算式を、数１７示す。

【０２３０】

【数１７】

【０２３１】Ｑｔを計算すると表５３のようになる。

【０２３２】

【表５３】

47　　48

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.00237	0.0028	0.0038	0.0054
－1.40	0.00231	0.00258	0.0034	0.0049
－1.40,-0.44	0.0031	0.00359	0.0051	0.0087

【０２３３】従って、－１．４０を外れ値とする。

【０２３４】実施例３２．検出統計量Ｉｔの計算式を、数１８示す。

【０２３５】

【数１８】

$$It = \frac{n+S}{(n-S)^2}\alpha$$

【０２３６】Ｉｔを計算すると表５４のようになる。

【０２３７】

【表５４】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.036	0.043	0.056	0.081
－1.40	0.035	0.039	0.051	0.074
－1.40,-0.44	0.046	0.054	0.077	0.130

【０２３８】従って、－１．４０を外れ値とする。

【０２３９】実施例３３．検出統計量Ｖｔの計算式を、数１９示す。

【数１９】

$$Vt = \frac{n-1}{(n-S)^2}\alpha$$

【０２４１】Ｖｔを計算すると表５５のようになる。

【表５５】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.033	0.037	0.045	0.063
－1.40	0.030	0.031	0.037	0.049
－1.40,-0.44	0.037	0.040	0.051	0.078

【０２４３】従って、－１．４０を外れ値とする。

【０２４４】実施例３４．検出統計量Ｅｔの計算式を、数２０示す。

【０２４５】

【数２０】

$$Et = \frac{n+1}{(n-S-1)^2}\alpha$$

【０２４６】Ｅｔを計算すると表５６のようになる。

【０２４７】

【表５６】

49　　50

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.038	0.043	0.055	0.075
－1.40	0.035	0.038	0.047	0.067
－1.40,-0.44	0.045	0.050	0.070	0.122

【0248】従って、－1．40を外れ値とする。

【0249】実施例35．検出統計量Ｊｔの計算式を、数21示す。

【0250】

【数21】

$$Jt = \frac{n(n+S+1)}{(n-S)^2}\alpha$$

【0251】Ｊｔを計算すると表57のようになる。

【0252】

【表57】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.569	0.634	0.782	1.040
－1.40	0.518	0.537	0.648	0.866
－1.40,-0.44	0.641	0.680	0.901	1.391

【0253】従って、－1．40を外れ値とする。

【0254】実施例36．検出統計量Ｖｔｄの計算式を、数22示す。

【0255】

【数22】

$$Vtd = n\log\alpha - \frac{b_2}{2} + 2*S$$

【0256】Ｖｔｄを計算すると表58のようになる。

【0257】

【表58】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	-11.57	-10.79	-8.90
－1.40	-12.54	-12.13	-10.55
－1.40,-0.44	-10.21	-9.79	-8.15

【0258】従って、－1．40を外れ値とする。

【0259】実施例37．検出統計量ＢＢｔの計算式を、数23示す。

【0260】

【数23】

$$BBt = \frac{n+1}{(n-S-1)^2}\alpha$$

【0261】ＢＢｔを計算すると表59のようになる。

【0262】

【表59】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.044	0.050	0.064
－1.40	0.041	0.044	0.056
－1.40,-0.44	0.052	0.059	0.084

【0263】従って、－1．40を外れ値とする。

【0264】実施例38．検出統計量ＣＣｔの計算式を、数24示す。

【0265】

【数24】

$$CCt = \frac{n(n-S+1)}{(n-S)^2}\alpha$$

【0266】ＣＣｔを計算すると表60のようになる。

【0267】

【表60】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.569	0.559	0.587
－1.40	0.453	0.403	0.405
－1.40,-0.44	0.481	0.431	0.450

【0268】従って、－1．40と1．01を外れ値とする。

51

【０２６９】実施例３９．検出統計量ＤＤｔの計算式を、数２５示す。

【０２７０】

【数２５】

$$DDt = \frac{n(n+S+1)}{(n-S)^2}\alpha$$

【０２７１】ＤＤｔを計算すると表６１のようになる。

【０２７２】

【表６１】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.589	0.634	0.782
－1.40	0.518	0.537	0.648
－1.40,-0.44	0.641	0.690	0.901

【０２７３】従って、－１．４０を外れ値とする。

【０２７４】実施例４０．検出統計量ＧＧｔの計算式を、数２６示す。

【０２７５】

【数２６】

$$GGt = \frac{n+S+1}{(n-S+1)^2}\alpha$$

【０２７６】ＧＧｔを計算すると表６２のようになる。

【０２７７】

【表６２】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.033	0.039	0.051
－1.40	0.032	0.035	0.044
－1.40,-0.44	0.042	0.047	0.063

52

＊【０２７８】従って、－１．４０を外れ値とする。

【０２７９】実施例４１．検出統計量Ｕｐｔの計算式を、数２７示す。

【０２８０】

【数２７】

$$Upt = \frac{n+S-1}{(n-S+1)^2}\alpha$$

【０２８１】Ｕｐｔを計算すると表６３のようになる。

【０２８２】

【表６３】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.487	0.478	0.531
－1.40	0.391	0.365	0.383
－1.40,-0.44	0.435	0.408	0.439

【０２８３】従って、－１．４０と１．０１を外れ値とする。

【０２８４】実施例４２．検出統計量Ｚｔの計算式を、数２８示す。

【０２８５】

【数２８】

$$Zt = \frac{n}{n-S}\alpha$$

【０２８６】Ｚｔを計算すると表６４のようになる。

【０２８７】

【表６４】

＊

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.534	0.515	0.538	0.585
－1.40	0.421	0.369	0.365	0.379
－1.40,-0.44	0.441	0.388	0.394	0.435

【０２８８】従って、－１．４０、１．０１、０．６３を外れ値とする。Ｚｔは外れ値を多めに検出する。

【０２８９】実施例４３．検出統計量Ｋｔの計算式を、数２９示す。

【０２９０】

【数２９】

$$Kt = \frac{1}{n-S}\alpha$$

【０２９１】Ｋｔを計算すると表６５のようになる。

【０２９２】

【表６５】

53　　54

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63	1.01 0.63 0.48
なし	0.036	0.037	0.041	0.049
－1.40	0.030	0.028	0.030	0.034
－1.40,-0.44	0.034	0.032	0.036	0.043

【０２９３】従って、－１．４０と１．０１を外れ値とする。

【０２９４】実施例４４．検出統計量Ｘｔの計算式を、数３０示す。

【０２９５】

【数３０】

$$Xt = \frac{1}{(n-S)(n+S)}\alpha$$

【０２９６】Ｘｔを計算すると表６６のようになる。

【０２９７】

【表６６】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	0.0024	0.0025	0.0028
－1.40	0.0020	0.0019	0.0020
－1.40,-0.44	0.0023	0.0022	0.0024

【０２９８】従って、－１．４０と１．０１を外れ値とする。

【０２９９】実施例４５．検出統計量ＨＱｔの計算式を、数３１示す。

【０３００】

【数３１】

$$HQt = n\log\alpha + C*S*\log(\log(n))$$
$$C>2$$

【０３０１】ＨＱｔを計算すると表６７のようになる。

【０３０２】

【表６７】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	-9.42	-7.41	-4.59
－1.40	-10.24	-9.47	-7.40
－1.40,-0.44	-7.16	-6.61	-4.72

【０３０３】従って、－１．４０を外れ値とする。

【０３０４】実施例４６．検出統計量ＡＩＣｔの計算式を、数３２示す。

【０３０５】

【数３２】

$$AICt = n\log\alpha^2 + 2*S$$

【０３０６】ＡＩＣｔを計算すると表６８のようになる。

【０３０７】

【表６８】

外れ値の 大きい方 候補 / 小さい方	なし	1.01	1.01 0.63
なし	-17.80	-17.61	-15.43
－1.40	-23.27	-25.19	-24.08
－1.40,-0.44	-20.59	-22.57	-21.39

【０３０８】従って、－１．４０と１．０１を外れ値とする。

【０３０９】

【発明の効果】第１の発明によれば、値を入力すれば算出された検出統計量に基づき外れ値が検出されるので、従来のように計算値と数表の大小比較をする必要がない。また、外れ値の個数を予め設定する必要がない。あるいは、外れ値として検出したい数の最大値を指定しておけばよい。また、外れ値の個数により、又は大きい方の外れ値が、小さい方の外れ値かにより計算方式を変える必要もないので、計算過程が簡単でかつ計算量も少なくてすむ。また、従来方式では、マスク効果により外れ値を検出できないことがあったが、これを回避することができるので、より正確な結果が得られる。また、外れ値が存在しない時は、存在しないと判定する。

【０３１０】第２の発明によれば、外れ値を算出するための値の選択が簡単に行える。

【０３１１】第３の発明によれば、検出統計量の単純な比較だけで外れ値を求めるので、処理が簡単になる。

【０３１２】第４の発明における計算式によれば、外れ値がある場合、最も外れた値が除かれると検出統計量は

55

最小となるので、これを利用して外れ値を求めることができる。

【０３１３】第５の発明における計算式によれば、補正項があるため、より的確に外れ値を求めることができる。

【０３１４】第６の発明における計算式によれば、第１の項目に分散を含んでいるため、最も外れている値を除くと分散が小さくなるという性質を利用できる。

【０３１５】第７の発明における計算式によれば、第２の項目に外れ値の候補の個数を含んでいるため、外れ値の個数を増やしていったことによる第１の項目の減少傾向を相殺できる。

【０３１６】第８の発明における計算式によれば、第２の項目に係数を乗算していることにより、第２の項目の増加量を調節することができる。

【０３１７】第９の発明における計算式によれば、分散の減少傾向を係数により、調節できる。

【０３１８】第１０の発明における計算式によれば、回帰分析の式を応用して外れ値の検出を行うことができる。

【０３１９】第１１の発明によれば、加工工程があることによりさまざまなタイプのデータの外れ値を検出することができる。

【０３２０】第１２の発明によれば、時間に依存する値からも時間に依存しない値に加工することにより、外れ値を検出することができる。

【０３２１】第１３の発明によれば、１つのサンプルに複数の特性値が存在する場合であっても、テコ比を計算することにより外れ値を求めることができる。

【０３２２】第１４の発明によれば、回帰分析の手法を適用できる値であれば、回帰分析の残差を求めることによりこの残差の外れ値を求めることができる。

【０３２３】第１５の発明によれば、正準相関分析モデルを適用できる場合でも、外れ値を求めることができる。

【０３２４】第１６の発明によれば、複数のグループに特性値が分類され、判別分析を行うことができる場合、外れ値を求めることができる。

【０３２５】第１７の発明によれば、上記のような外れ値検出方法を利用することにより、外れ値を容易に検出することができるデータ処理装置を得ることができる。この外れ値が得られた時の環境条件を検討することにより、新たな知見・情報が得られることにもなる。

56

【０３２６】第１８の発明によれば、上記のような外れ値検出方法を利用することにより、外れ値を容易に検出し、除くことができるデータ処理装置を得ることができる。このデータ処理装置により得られた結果は、外れ値を除いてあるので信頼性が向上している。

【図面の簡単な説明】

【図１】本発明の外れ値検出方法を説明するためのフローチャート図である。

【図２】Ｇｒｕｂｂｓのデータ１を用いた場合の外れ値の候補の個数ｓと検出統計量Ｕｔの関係を示す図である。

【図３】別のデータを用いた場合の外れ値の候補の個数ｓと検出統計量Ｕｔの関係を示す図である。

【図４】本発明のデータ処理装置の構成図である。

【図５】本発明の一実施例の入力データをプロットした図である。

【図６】本発明の外れ値検出のための工程を説明する図である。

【図７】本発明の一実施例の時間とともに増加する傾向を持つデータをプロットした図である。

【図８】本発明の一実施例の時間とともに増加する傾向を除いたデータをプロットした図である。

【図９】本発明の一実施例の入力データをプロットした図である。

【図１０】回帰分析説明変数選択基準の数式を示す図である。

【図１１】本発明の実施例の中で使われるデータをプロットした図である。

【図１２】従来の技術及び本発明の実施例の中で使われるデータをプロットした図である。

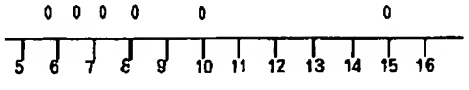
【図１３】従来の技術及び本発明の実施例で使われる装置の構成図である。

【図１４】従来の外れ値検出方式を説明するためのフローチャート図である。

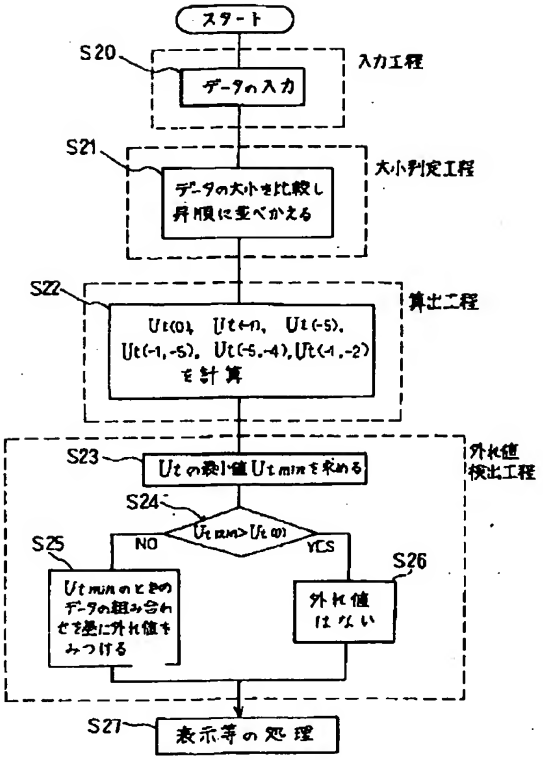
【符号の説明】

1　情報処理装置
2　コンピュータ（ＦＤＤ付き）
3　ディスプレイ・ユニット
4　プリンタ
5　キーボード
6　フロッピーディスク

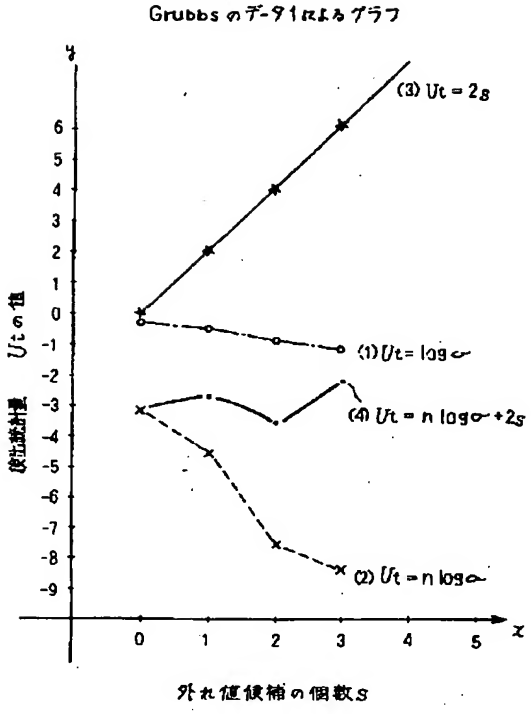
【図１２】



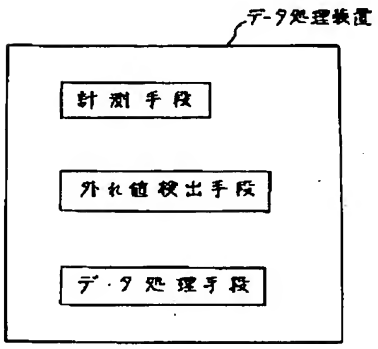
【図1】


スタート
S20
入力工程
データの入力
S21
大小判定工程
データの大小を比較し昇順に並べかえる
S22
算出工程
Ut(0)、Ut(-1)、Ut(-5)、Ut(-1,-5)、Ut(-5,-4)、Ut(-1,-2)を計算
外れ値検出工程
S23
Utの最小値Ut minを求める
S24
Ut min>Ut(0)
NO
YES
S25
Ut minのときのデータの組み合わせを基に外れ値をみつける
S26
外れ値はない
S27
表示等の処理

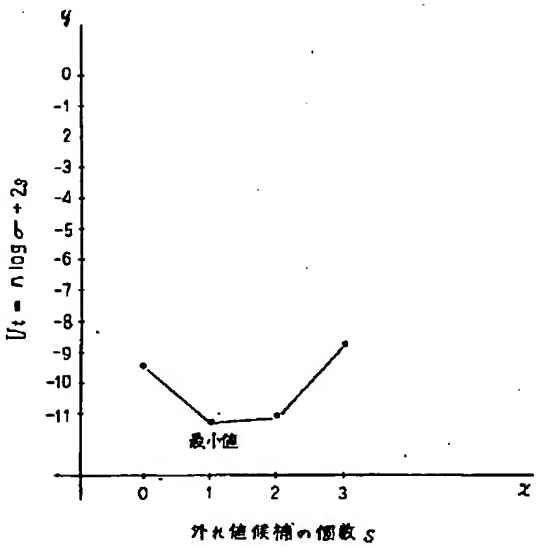
【図2】


Grubbsのデータによるグラフ
y
x
Utの値
検出統計量
(3) Ut = 2s
(1) Ut = log σ
(4) Ut = n log σ + 2s
(2) Ut = n log σ
外れ値候補の個数s

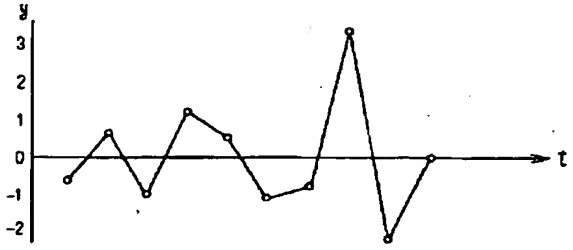
【図4】


データ処理装置
計測手段
外れ値検出手段
データ処理手段

【図3】

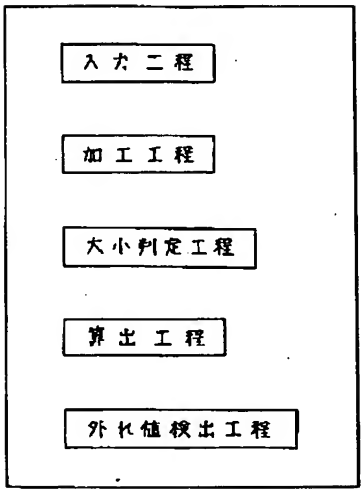

y
x
Ut = n log σ + 2s
最小値
外れ値候補の個数s

【図8】

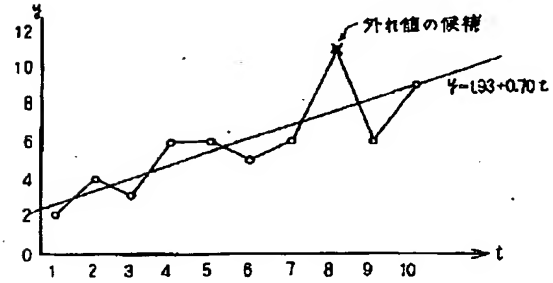

y
t

【図5】

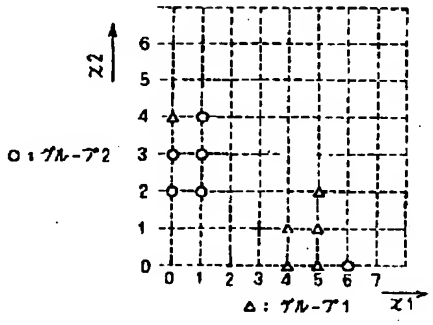
【図6】



【図7】

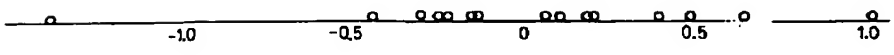


【図9】



【図11】

実施例18～46のデータのプロット

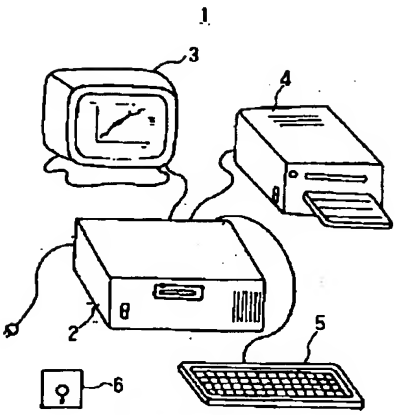


【図１０】

回帰分析説明変数選択規準

1	aic	$= n \log(\Sigma e_i^2) + 2P$	n：データの個数
2	uc	$= aic + P(P-2)/n$	P：説明変数の個数
3	姓和	$= 1 - \frac{(n-2)(n-1)}{(n-P-2)(n-P-1)}(1-R^2)$	R：重相関係数
4	芳賀他	$= 1 - \frac{(n-1)(n+P+1)}{(n+1)(n-P-1)}(1-R^2)$	
5	竹内	$= \frac{n^2-n-P-2}{n(n-P-2)(n-P-1)}\sigma^2$	σ^2：残差の分散
6	予測	$= \frac{(n+1)(n-2)}{n(n-P-2)}\sigma^2$	
7	Sp	$= \frac{\Sigma e_i^2}{(n-P-1)(n-P-2)}$	Σe^2：残差の2乗和
8	上田	$= 1 - \frac{(n+P+1)}{(n-P-1)}(1-R^2)$	
9	aicc	$= n \log(\Sigma e_i^2) + C \cdot P$	$C > 0$
10	HQ	$= n \log(\Sigma e_i^2) + C \cdot P \cdot \log(\log(n))$	$C > 2$
11	fpe	$= \frac{n+P}{n-P}\Sigma e_i^2$	
12	Helms	$= \frac{P}{n(n-P)}\Sigma e_i^2$	
13	シュワルツ	$= P \log n + n \log \sigma$	

【図１３】



1：情報処理装置
2：コンピュータ(FDD付き)
3：ディスプレイ
4：プリンタ
5：キーボード
6：フロッピーディスク

【図１４】